

Open Research Online

The Open University's repository of research publications and other research outputs

A content-aware quantisation mechanism for transform domain distributed video coding

Conference or Workshop Item

How to cite:

Mahmood, Asif; Dooley, Laurence S. and Wong, Patrick (2018). A content-aware quantisation mechanism for transform domain distributed video coding. In: The 2017 International Conference on Computational Science and Computational Intelligence, 14-16 Dec 2017, Las Vegas, USA.

For guidance on citations see [FAQs](#).

© [not recorded]



<https://creativecommons.org/licenses/by-nc-nd/4.0/>

Version: Accepted Manuscript

Link(s) to article on publisher's website:
<http://dx.doi.org/doi:10.1109/CSCI.2017.93>

Copyright and Moral Rights for the articles on this site are retained by the individual authors and/or other copyright owners. For more information on Open Research Online's data [policy](#) on reuse of materials please consult the policies page.

oro.open.ac.uk

A Content-Aware Quantisation Mechanism for Transform Domain Distributed Video Coding

Asif Mahmood, Laurence S. Dooley, Patrick Wong

School of Computing and Communications
The Open University
Milton Keynes, UK

Email: asif.mahmood@open.ac.uk, laurence.dooley@open.ac.uk, patrick.wong@open.ac.uk

Abstract—The *discrete cosine transform* (DCT) is widely applied in modern codecs to remove spatial redundancies, with the resulting DCT coefficients being quantised to achieve compression as well as bit-rate control. In *distributed video coding* (DVC) architectures like DISCOVER, DCT coefficient quantisation is traditionally performed using predetermined *quantisation matrices* (QM), which means the compression is heavily dependent on the sequence being coded. This makes bit-rate control challenging, with the situation exacerbated in the coding of high resolution sequences due to QM scarcity and the non-uniform bit-rate gaps between them. This paper introduces a novel *content-aware quantisation* (CAQ) mechanism to overcome the limitations of existing quantisation methods in transform domain DVC. CAQ creates a frame-specific QM to reduce quantisation errors by analysing the distribution of DCT coefficients. In contrast to the predetermined QM that is applicable to only 4x4 block sizes, CAQ produces QM for larger block sizes to enhance compression at higher resolutions. This provides superior bit-rate control and better output quality by seeking to fully exploit the available bandwidth, which is especially beneficial in bandwidth constrained scenarios. In addition, CAQ generates superior perceptual results by innovatively applying different weightings to the DCT coefficients to reflect the human visual system. Experimental results corroborate that CAQ both quantitatively and qualitatively provides enhanced output quality in bandwidth limited scenarios, by consistently utilising over 90% of available bandwidth.

Keywords—*distributed video coding; Wyner-Ziv video coding; content-aware quantisation; adaptive quantisation; bit allocation;*

I. INTRODUCTION

Distributed Video Coding (DVC) [1] is an alternate coding paradigm which compresses video content by exploiting statistical data redundancies at the decoder. In contrast to conventional codecs [2], [3], the DVC encoder design is thus much simpler, though the corresponding decoder design is more complicated. This complexity redeployment is attractive for many applications such as low-power surveillance networks, mobile video cameras and for resource-scarce Internet of Things devices.

The foundations of DVC are the information theoretic results of Slepian-Wolf [4] and Wyner-Ziv [5]. The *Slepian-Wolf* (SW) theorem deals with lossless compression of multiple-correlated sources, while the *Wyner-Ziv* (WZ)

theorem considers the lossy compression equivalent using so-called *side-information* (SI) at the decoder. Both theorems prove that no rate loss is incurred provided the source statistics are exploited only at the decoder compared to the orthodox approach of undertaking such processing at the encoder. The corollary is that complexity and the corresponding resource requirements for DVC encoders are appreciably relaxed compared to their conventional counterparts which do not exploit any statistical dependencies. Since the inception of DVC, extensive research have been conducted and several architectures proposed to realise the SW and WZ theoretical results in practice. The Stanford architecture [1] has become widely adopted by the community chiefly because of its extendible and scalable design. Among its successors, the DISCOVER codec [6] enhanced the Stanford architecture by relaxing some assumptions and introducing more sophisticated coding techniques, to the extent it has become the basis of most state-of-the-art DVC solutions [7], [8]. This has helped establish the superiority of DVC in terms of *rate distortion* (RD) performance over conventional low complexity codecs like H.264/AVC Intra at low spatial resolutions. However it has yet to surpass conventional complex inter-frame codecs and is significantly restricted by poor coding performance for higher resolution sequences [9].

Transform domain DVC [10] achieves better RD performance than its pixel-based counterpart [11] by exploiting spatial redundancies however, little research has been undertaken into the underlying transformation and quantisation processes. Most *transform domain Wyner-Ziv* (TDWZ) video coding architectures use the *discrete cosine transform* (DCT) and 4x4 block sizes, although the discrete wavelet transform has been proposed as a transform alternative in [12]. In the DCT-based TDWZ coding architecture, seven *quantisation matrices* (QM) [10] are used to facilitate limited control over the output quality and bit-rate, with an additional QM included for very high bit-rates in [13]. These eight QM have been empirically determined from a raft of test sequences and have been adopted in subsequent TDWZ architectures, though while these QM provide a useful set for encoding experiments, they are inadequate for many practical applications. They are also only applicable to 4x4 block-sized DCT, and while this may be pragmatic at low spatial resolutions (QCIF), it is often impractical at higher resolutions. This observation is substantiated in [14] where larger DCT block sizes have

been shown to perform better than 4x4 blocks at higher resolutions, with the codec design for the state-of-the-art H.265/HEVC [2] allowing up to 32x32 transform block sizes.

The main drawback of using a predetermined QM is its lack of flexibility in rate control and inefficient bandwidth utilisation. In DISCOVER [6] for example, a QM from the set of eight is chosen during initialisation and then used throughout the encoding of a sequence. As compression performance can vary significantly depending on SI quality, bit-rates for different sequences coded using the same QM can differ markedly, so making efficient bandwidth utilisation problematic in resource-constrained scenarios. Considerable available bandwidth may be idle when coding low motion sequences, while detrimental latency and visual artefacts due to bandwidth excess may be observed in high motion sequences. This problem is compounded when encoding sequences where either the SI quality varies due to changes in the scene characteristics or in variable bandwidth limit scenarios.

Switching QM during coding is not viable because the QM are non-uniformly spaced in terms of bit-rates. This becomes more problematic when encoding higher resolution sequences as changing the QM can lead to significant bit-rate variations. Some *region of interest* (ROI) based coding schemes [15] attempt to control the bit-rate by classifying the blocks in a WZ frame into multiple categories and then applying a different QM to each category. This is not however, a generic solution because firstly, there is no universal ROI selection method that can be applied to each scenario and secondly, ROI methods usually increase encoder complexity and transmission overheads. Moreover, ROI coding may lead to visual artefacts and overall inferior coding quality of areas lying outside the ROI.

To address these issues, this paper introduces a novel quantisation mechanism for TDWZ video coding called *content-aware quantisation* (CAQ) which dynamically derives the QM for a target bandwidth. CAQ is incorporated into the DVC decoder and generates a customised QM for each WZ frame by analysing the coefficient distribution of the corresponding SI to maximise the bandwidth utilisation and thereby give superior rate control. It can readily adapt to changes in scene characteristics as well as bandwidth limits. Importantly, CAQ does not depend on the transform block size so it can be employed to produce QM for larger block sizes, which is advantageous in encoding high resolution. The impact of larger DCT block sizes using CAQ have been analysed in [14] so the focus in this paper is upon the CAQ design mechanism and a critical evaluation of its bandwidth utilisation performance compared to existing predetermined QM for 4x4 block-sizes. In addition, the perceptual sensitivity of the human eye to different DCT coefficients is exploited in CAQ by integrating a weighting scheme which is shown to improve the qualitative output quality for a range of bandwidth limit scenarios compared with the predetermined QM method. The numerical results analysis corroborates the superiority of CAQ in terms of consistently better utilisation of the available bandwidth capacity. It also exhibits improved RD performance in comparison to

DISCOVER for sequences with low-motion activity and equivalent RD performance for all other sequence types.

The remainder of the paper is organised as follows: Section II provides an overview of the quantisation process of TDWZ coding architectures, while Section III details the new CAQ mechanism. Section IV presents a critical comparative performance analysis, with Section V providing some concluding comments.

II. BACKGROUND

The quantisation mechanism of the DISCOVER DVC codec [6] will firstly be reviewed as a representative of TDWZ coding architectures, because it is an effective and practical TDWZ solution. It is employed as the basis of many advanced TDWZ solutions [9], [16], serving as a benchmark for evaluating DVC codec performance, and for this reason it is adopted as the test development platform in this paper, since improvements in this architecture will be transferable to more advanced codec variants.

In DISCOVER, the encoder divides the input frames of a video sequence into *key frames* (KF) and WZ frames according to *group of pictures* (GOP) configuration. GOP 2 denotes alternating KF and WZ frames, while GOP 4 refers to each KF being followed by three WZ frames. The KF are transmitted to the decoder using a conventional low complexity H.264/AVC Intra codec, while WZ frames are DVC coded, with each frame firstly transformed using a block-based DCT followed by quantisation using a predetermined QM. A *low-density parity check accumulate* (LDPCA) encoder [17] then generates error correcting information (parity bits) from quantised coefficients and transmits them to the decoder, while actual coefficients are ignored. At the decoder, the SI is constructed from neighbouring KF to generate an estimate of the WZ frame, which is then in turn similarly transformed and quantised. The quantised SI coefficients act as noisy transmitted data at the LDPCA decoder which performs error correction using the information received from the encoder. In order to facilitate rate control and compression, only a small number of parity bits are initially transmitted while the remainder are buffered at the encoder. The channel decoder attempts to remove noise using the currently available parity bits and if this is ineffective, it requests more parity bits via a feedback channel. After successful recovery, the coefficients are reconstructed from the recovered quantised symbols and SI, with an inverse DCT being applied to generate the decoded WZ frame.

DISCOVER employs a 4x4 block based DCT to remove WZ frame redundancies together with a fixed QM [10], [13] for quantisation. Since the LDPCA coder performs better on longer sequences, the transformed coefficients from different blocks corresponding to the same DCT frequency component are grouped together to form bands. Each coefficient band is collectively quantised, with the number of steps derived from the QM. Each quantised coefficient band is then converted into bit planes before being processed by the LDPCA coder. To achieve maximum bandwidth utilisation, the number of quantisation bins is always a power of 2, so the QM used by DISCOVER essentially defines the number of bit-planes to

be extracted from each coefficient band. It also constrains the design of any other QM because quantising with an arbitrary number of bins leads to bandwidth waste.

While the DCT is fully reversible, to reduce the bit-rate, the coefficients are quantised which may introduce visual distortion. Since the human eye is more sensitive to lower frequency components, higher frequencies are more heavily quantised in DISCOVER to achieve better compression. However, since coefficient distribution is image dependent, collectively quantising higher frequency coefficients can introduce significant perceptual distortion in the decoded frame. Content-adaptive quantisation methods [18] potentially offer superior output qualities compared to the traditional static approach by allocating finer quantisation to more prominent coefficient bands. This is however, intractable to achieve in practical bandwidth-constrained scenarios because allocating a fine quantiser with more bins to one coefficient band, reduces the number of available bins for other bands.

Bit-rates for DISCOVER primarily depend on the quality of SI, so using the same QM to code sequences with disparate scene characteristics can result in significantly difference in the resulting bit-rates. Dynamically switching QM sets is unreliable since the number of bit-planes and resulting bandwidth is not uniform across the QM as illustrated in Table I. For example, the bandwidth requirement for QM₄ is almost twice that of QM₃, while switching to QM₆ from QM₇ saves only 10% of the bandwidth, assuming equi-compression of the bit-planes. It is thus challenging to adaptively change QM according to scene characteristics and bandwidth. In contrast, this paper presents a novel solution to this problem by incorporating a CAQ mechanism which generates a customised QM tailored to a sequence so as to maximise the bandwidth utilisation. The new CAQ algorithm is presented in the next section.

TABLE I. NUMBER OF BIT-PLANES PRODUCED BY QM IN DISCOVER

QM	1	2	3	4	5	6	7	8
Bit-planes	10	11	16	30	36	45	50	63

III. CONTENT-AWARE QUANTISATION (CAQ)

Fig. 1 shows the block diagram of DISCOVER DVC architecture along with the new CAQ module incorporated at the decoder. CAQ takes the number of *available bit-planes* (ABP) as input and analyses the distribution of SI coefficients. It then derives a custom QM which minimises the overall magnitude of quantisation errors for all coefficient bands. The custom QM is transmitted back to the encoder via the feedback channel which is then used for quantisation.

CAQ firstly determines the number of ABP to be distributed among the coefficient bands. In a bandwidth constraint scenario, this can be derived from the available bandwidth and correlation noise model information in a similar manner to predict the required bit-rate at the encoder of a TDWZ coder [19], [20]. Alternatively, the ABP value can be supplied by the decoder and used by the CAQ module

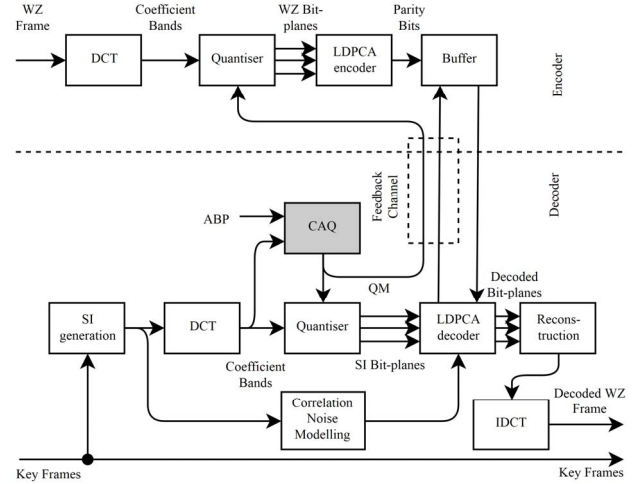


Figure 1. Block diagram of the DISCOVER codec architecture with the new CAQ module highlighted

to produce a custom QM for the WZ frames in order to ensure the desired output quality. This design permits a new ABP value for every WZ frame so facilitating greater flexibility in controlling both the bit-rate and decoded output quality.

Once the ABP value has been determined, the CAQ algorithm starts allocating bit planes to the coefficient bands. This allocation is undertaken to minimise the most prominent quantisation errors. The coefficient values are reconstructed from the decoded quantised symbols and SI [21], however instead of scaling as in conventional codecs. Because the exact quantisation error is not known, an estimate of the error for each coefficient is framed as the absolute difference between the coefficient value and the quantised symbol, multiplied by the step size. CAQ is initialised with an empty QM and zero bit-planes are allocated. It then estimates the total quantisation error for each coefficient band before allocating a bit-plane to the band with highest total. This recursive process continues until all the bit-planes have been allocated.

To illustrate the CAQ mechanism, consider the scenario where N coefficient bands are formed from the DCT coefficients of the SI, with each band containing L coefficients. N depends on the DCT block size, while L can be derived from N and the spatial resolution of the sequence. So, to encode a *common interchange format* (CIF) resolution video (352x288 pixels), 4x4 DCT block sizes are used, forming 16 coefficient bands, each containing 6336 coefficients i.e., $N = 16$ and $L = 6336$. More generally, if c_{ij} is a DCT coefficient in coefficient band j of the SI and M_j is the corresponding number of bit-planes for this band, then j will have values lying in the range $[1, N]$ and the corresponding range for i will be $[1, L]$. QM can now be framed in terms of the number of bit-planes allocated to each coefficient band, i.e. $QM = \{M_1, M_2, \dots, M_N\}$.

If q_{ij} and e_{ij} are respectively the quantised coefficient corresponding to c_{ij} and the estimated quantisation error, then these are expressed as:

$$q_{ij} = \text{round}\left(\frac{c_{ij}}{\Delta_j}\right) \quad (1)$$

$$e_{ij} = |c_{ij} - q_{ij}\Delta_j| \quad (2)$$

where $\text{round}(a)$ is the nearest integer of a and Δ_j denotes the step size of the quantiser, which is calculated from the dynamic range of the band j and M_j . If the maximum absolute value in each coefficient band is MaxVal_j , then the step size Δ_j is given by:

$$\Delta_j = \left\lceil \frac{2 \times \text{MaxVal}_j}{2^{M_j}} \right\rceil \quad (3)$$

so the estimated quantisation error E_j for each band j is then given by the total of individual estimated errors of all coefficients within the band.

$$E_j = \sum_{i=1}^L e_{ij} \quad (4)$$

CAQ is initialised with an empty QM, i.e. $M_j = 0$ for $j \in \{1, 2, \dots, N\}$. It calculates Δ_j and E_j for all coefficient bands to identify the band b with the largest estimated quantisation error, which is then allocated a bit-plane:

$$b = \arg \max_{1 \leq j \leq N} E_j$$

$$M_b = M_b + 1 \quad (5)$$

After allocation, Δ_b is updated using (3) and q_{ib} , e_{ib} and E_b are recomputed for band b with the new Δ_b in (1), (2) and (4) respectively. This process is repeated to find the largest E_b and the next bit-plane allocated to the corresponding coefficient band, and continues until all available bit-planes have been allocated.

When a bit-plane is allocated to a coefficient band, the step size of corresponding quantiser is changed according to (3). The new step size becomes approximately half of the previous and since the quantisation error is limited by the step size, the maximum error for that band is similarly halved. Thus, by allocating a bit-plane to the band with largest quantisation error, CAQ reduces both the quantisation errors and resulting distortions in the decoded DVC output.

The complete CAQ algorithm for progressively allocating available bit-planes is presented in pseudo-code format in Algorithm 1.

Steps 1–9 represent the initialisation stage using an empty QM and a priori quantisation error values. The coefficient band with the largest quantisation error is then identified and a bit-plane allocated to that band (Steps 11–12) using (5). Both the quantiser step size and quantisation error corresponding to the band are updated (Steps 13–18) and the process repeated (Steps 10–19) until all available bit-planes have been assigned.

Algorithm 1 Customised QM from a given number of available bit-planes

Input: ABP // Number of available bit-planes
Output: QM // Quantisation matrix

```

1: for  $j = 1$  to  $N$  do
2:    $M_j \leftarrow 0$ 
3:   Calculate  $\Delta_j$  using (3)
4:   for  $i = 1$  to  $L$  do
5:     Calculate  $q_{ij}$  using (1)
6:     Calculate  $e_{ij}$  using (2)
7:   end for
8:   Calculate  $E_j$  using (4)
9: end for
10: for  $a = 1$  to  $ABP$  do
11:   Find  $b$  such that  $E_b = \max(E_j)$ 
12:    $M_b \leftarrow M_b + 1$ 
13:   Update  $\Delta_b$  using (3)
14:   for  $i = 1$  to  $L$  do
15:     Calculate  $q_{ib}$  using (1)
16:     Calculate  $e_{ib}$  using (2)
17:   end for
18:   Update  $E_b$  using (4)
19: end for

```

A. Integrating the human visual system (HVS) into CAQ

Algorithm 1 treats all DCT coefficient bands equally and minimises the overall magnitude of the quantisation errors in the transform domain. However, the eye is more sensitive to errors at lower frequencies than higher frequencies, so by giving equal weight to all coefficient bands, CAQ may allocate some bit-planes to higher frequency bands that have a larger error than a lower frequency component. It is thus possible for coefficient bands to have lower quantisation error magnitudes, while the corresponding decoded output has more visible artefacts.

This uneven sensitivity reflects the way the *human visual system* (HVS) operates and so in order to make the DCT coefficients perceptually equal, a set of weights are applied to the different DCT frequency components, in an analogous manner to the quantisation tables (Q-tables) employed in JPEG [22]. JPEG however, uses 8x8 DCT block sizes so these QM are not applicable to the 4x4 block sizes used in TDWZ coding. To incorporate the HVS into CAQ, a customised weighting matrix for 4x4 block sizes has been derived based on [23], and this is displayed in Fig. 2.

All coefficient bands are pre-processed to accommodate HVS in CAQ, by weighting each band by the corresponding weight in Fig. 2 prior to applying Algorithm 1. The resulting

1.0000	1.0000	0.9599	0.7684
1.0000	0.9571	0.8192	0.6471
0.9599	0.8192	0.5419	0.3930
0.7684	0.6471	0.3930	0.2278

Figure 2. HVS weight matrix for 4x4 DCT block size

custom QM is then transmitted to the encoder via the feedback channel, incurring a very small resource cost. For 8 bits per pixel sequences, the DCT coefficients, after rounding to the nearest integer, are represented by 11 bits, so up to 11 bit-planes may be generated in any band i.e., a 4 bit overhead. This means for 4x4 DCT block sizes, the feedback channel expends just 64 bits per frame. An important point to stress is while CAQ produces a customised QM based on analysing the distribution of weighted coefficients, the actual quantisation is performed on the unweighted coefficients in exactly the same way as in the fixed QM approach. This means the decoder has the flexibility to apply any weighting scheme germane to an application or domain. The performance of the HVS weighted CAQ mechanism will now be investigated.

IV. EXPERIMENTAL RESULTS DISCUSSION

A series of experiments were conducted for the new weighted CAQ mechanism by seamlessly integrating it into the DISCOVER codec architecture. This modified DISCOVER codec including CAQ, was implemented in C++ and used OpenCV library routines, with the simulation computing platform being an Intel Core i7 processor running Windows 7.

To critically evaluate the comparative CAQ performance, a set of diverse test sequences has been encoded by both the modified DISCOVER including CAQ, and the original DISCOVER codec. The test sequences exhibit a variety of challenging features including different types of motion, multiple objects, textures and occlusions and while QCIF (176x144 pixels) resolution is commonly used for DVC testing, higher resolution CIF (352x288 pixels) sequences have instead been applied to appraise CAQ robustness. All sequences used GOP size of 2, i.e. alternate KF and WZ frames, and the corresponding bit-rate and *peak signal to noise ratio* (PSNR) performance measured. In addition, since CAQ has no RD performance impact on KF, only results for different quantisation methods for the WZ frames are analysed in this discussion.

Fig. 3 displays the comparative performance of the custom CAQ-generated QM (denoted as CAQ) and the fixed QM of DISCOVER (denoted as DISCOVER) for Football in terms of flexible rate control, bandwidth utilisation and PSNR improvement. All the other elements of the encoding/decoding processes for DISCOVER are kept identical to ensure an equitable comparison. The results show the RD points for the first 7 fixed QM of DISCOVER, with QM₈ being omitted because it incurs an unrealistically high bit-rate of 3084.368kbps. In contrast, 34 RD points for CAQ are shown which have been derived using ABP values in the range [16, 49]. These are evidently more regularly spaced in terms of bit-rate compared to those of the fixed QM, which implies much greater rate control flexibility thus enabling more effective bandwidth utilisation and corresponding PSNR gains. For example, for a bandwidth limit of 1600kbps, DISCOVER achieves the best output using QM₆, utilising 73% of the available bandwidth. In contrast, by incorporating CAQ, the best achievable output is obtained for an ABP value of 47, utilising over 98% of the

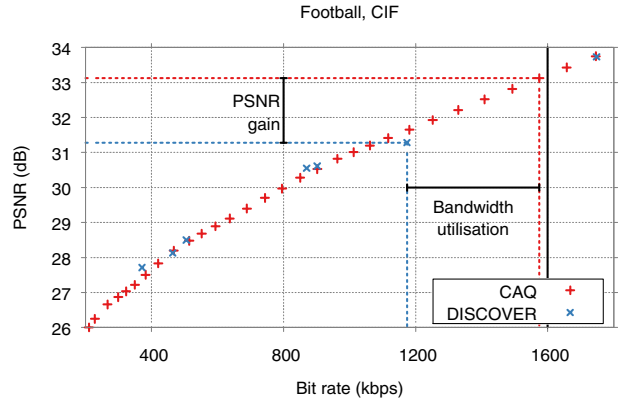


Figure 3. Rate control flexibility and resulting RD performance improvement of CAQ compared to fixed QM of DISCOVER.

bandwidth and giving a PSNR gain of 1.84dB over DISCOVER. Furthermore, the decoder can choose up to 6 ABP values between 42 and 47, each of which utilises the bandwidth more effectively and produces better output quality, thereby corroborating the superior flexible rate control performance of CAQ.

The corresponding RD performance of fixed QM at low bit-rates is slightly better than the RD performance of CAQ in certain cases, however CAQ is still advantageous because it produces outputs at very low bit-rates which the fixed QM cannot. For example, the minimum output quality produced by DISCOVER is 371.263kbps so if the available bandwidth falls below this value, DISCOVER does not produce an output. In contrast, CAQ not only generates an output, but affords multiple RD selection points to match a prescribed bandwidth requirement.

Table II summarises the bandwidth utilisation results for various test sequences encoded by two different quantisation mechanisms namely: i) DISCOVER with the custom QM from HVS weighted CAQ and ii) DISCOVER with predetermined QM. In addition, several bandwidth limits have been specified for the experiments, with in each case, the maximum achievable PSNR obtained by both methods being presented alongside the corresponding bit-rate represented as a percentage relative to the hypothetical bandwidth limit.

The Table II results consistently reveal superior bit-rate control using CAQ compared to the fixed QM method in DISCOVER, with more than 90% of the available bandwidth capacity utilised in all test sequences, allied with improved PSNR performance. Conversely, bandwidth utilisation for DISCOVER is inconsistent, so while it uses $\approx 99\%$ of the 200kbps capacity for Hall, this drops to only 62% for the 400kbps bound for Coastguard. Moreover, the RD results for CAQ have been produced using a fixed ABP value for all frames within a sequence, so potentially further improvements in utilisation are feasible by dynamically determining ABP from the bandwidth capacity and noise model information.

In addition to effective bandwidth utilisation, another contributory factor to PSNR improvement is the intelligent

TABLE II. BANDWIDTH UTILISATION STATISTICS FOR VARIOUS QUANTISATION METHODS

Bandwidth Limit (kbps)		200		400		800		1600	
Sequence	Quantisation Method	Bit-rate (%)	PSNR (dB)	Bit-rate (%)	PSNR (dB)	Bit-rate (%)	PSNR (dB)	Bit-rate (%)	PSNR (dB)
Football	CAQ	91.48	25.66	95.55	27.50	99.43	29.97	98.41	33.12
	DISCOVER			92.82	27.71	63.13	28.50	73.36	31.28
Soccer	CAQ	95.83	29.95	99.30	31.96	94.35	34.14	99.30	37.12
	DISCOVER			92.10	31.95	93.27	34.13	93.39	36.79
Foreman	CAQ	98.67	33.97	96.25	35.02	95.47	36.46	98.87	38.45
	DISCOVER	87.77	34.10	67.35	34.74	89.53	36.19	67.19	37.21
Coastguard	CAQ	98.15	33.09	92.32	33.95	98.33	35.29	98.60	37.16
	DISCOVER	78.85	33.11	62.05	33.70	96.17	35.14	74.57	36.27
Mother & Daughter	CAQ	99.63	42.51	99.86	43.42	98.22	44.54	98.66	46.00
	DISCOVER	81.67	42.51	84.92	42.90	95.04	44.09	98.88	45.64
Hall	CAQ	99.88	36.88	98.37	37.86	93.87	39.18	97.60	40.88
	DISCOVER	99.40	37.12	92.57	37.80	90.54	39.10	83.30	40.47

CAQ bit-plane allocation which seeks to minimise the quantisation errors. This is evident in the comparative results for Mother & Daughter at the 1600kbps limit, where CAQ produces better PSNR values at a lower bandwidth than DISCOVER. The performance of CAQ does tend to slightly degrade at very low bit-rates, because it generates a custom QM based on SI coefficient distribution, which may be an inadequate approximation of the original WZ frame. As there are very few bit-planes available at such low bit-rates, any misallocation can impact the overall output quality considerably leading to slight performance degradations. Despite this seeming restriction, CAQ can still be advantageously applied in low bandwidth cases because it is able to generate outputs where DISCOVER may fail. In both Football and Soccer for instance, the lowest rate QM₁ of DISCOVER requires a bit-rate higher than the available bandwidth capacity of 200kbps, so no output is generated, while CAQ is still able to derive a valid QM at this capacity value.

To complement the CAQ numerical analysis, a qualitative assessment was also undertaken to corroborate its enhanced performance in bandwidth constrained scenarios. Fig. 4 shows an example frame (#42) of Football using both quantisation mechanisms at the 1600kbps limit, with the respective QM for each frame included. It is visually apparent that the CAQ output has significantly lower perceptual artefacts than decoded DISCOVER output, a trend observed in other test sequences.

Since CAQ generates the QM from SI which is transmitted back to the encoder, a small delay is incurred, though in terms of complexity, CAQ is computationally significantly more efficient compared to either the SI generation process or the iterative decoding algorithm of LDPCA coder, so any time overhead incurred will be negligible. If latency is a design priority, the custom QM can instead be generated from the previous KF rather than SI to reduce the feedback delay though this may lead to reduced performance if there are sudden scene changes.

In summarising, both quantitative and qualitative results confirm the key advantage of CAQ is flexible rate control, which is especially beneficial in bandwidth limited scenarios where it can more reliably utilise the available bandwidth

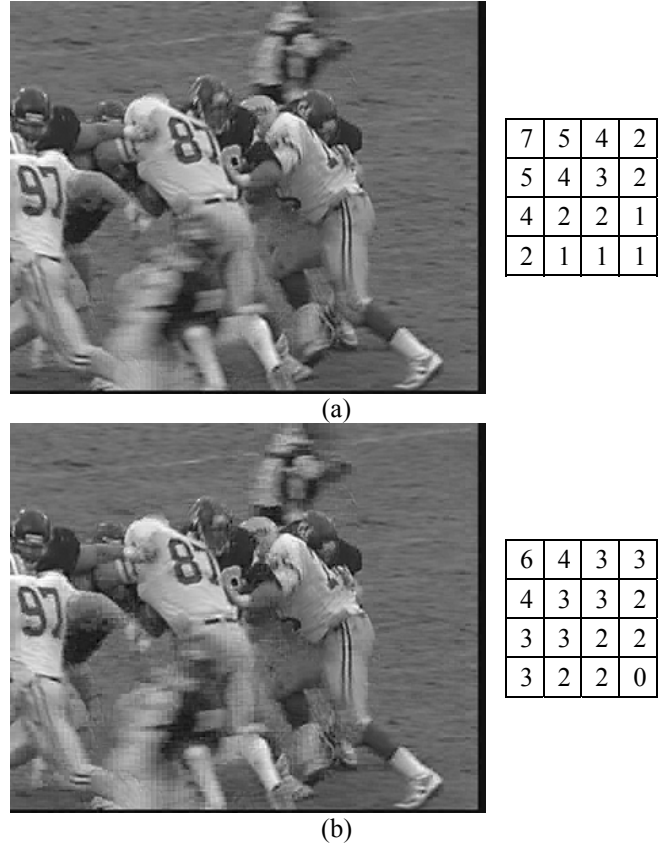


Figure 4. Perceptual comparison of frame #42 for Football decoded using different quantisation mechanisms (a) CAQ and (b) DISCOVER along with their corresponding QM.

leading to superior decoded output quality. It consistently outperforms quantisation using a predetermined QM in low-motion sequences where the decoder produces high SI quality. It also exhibits comparable RD performance for high motion sequences which have lower SI quality. Applying more sophisticated SI generation techniques [7], [8], [24] will further enhance CAQ performance and while a low bit-rate feedback channel is required, CAQ crucially can produce a QM for any transform block size, so it can be

exploited in coding larger block sizes. Since CAQ is incorporated as a new module at the DISCOVER decoder and does not replace any existing module, it can be embedded into any current state of the art transform domain DVC architecture based upon DISCOVER [9], [16] to improve its rate-control performance in both mono and multiview scenarios.

V. CONCLUSION

This paper has presented a novel *content-aware quantisation* (CAQ) mechanism for transform domain DVC. Quantisation of the transform coefficients is essential for compression so using a predetermined *quantisation matrix* (QM) introduces significant constraints upon both bandwidth utilisation and bit-rate control. CAQ ensures superior utilisation of the available bandwidth in resource constrained scenarios by progressively allocating bit-planes to different coefficient bands. It also enhances the quality of the decoded DVC output by integrating the human visual system into CAQ in order to apply appropriate weights to the coefficient bands, with the decoded output having less perceptible distortion. Experimental results confirm the superior performance of CAQ for CIF resolution sequences, while comparing the respective RD performance, CAQ outperforms the predetermined QM approach of DISCOVER for low motion sequences and affords competitive performance for other sequence types.

REFERENCES

- [1] B. Girod, A. M. Aaron, S. Rane, and D. Rebollo-Monedero, 'Distributed video coding', *Proceedings of the IEEE*, vol. 93, no. 1, pp. 71–83, Jan. 2005.
- [2] G. J. Sullivan, J.-R. Ohm, W.-J. Han, and T. Wiegand, 'Overview of the high efficiency video coding (HEVC) standard', *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 22, no. 12, pp. 1649–1668, 2012.
- [3] T. Wiegand, G. J. Sullivan, G. Bjontegaard, and A. Luthra, 'Overview of the H.264/AVC video coding standard', *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 13, no. 7, pp. 560–576, Jul. 2003.
- [4] D. Slepian and J. K. Wolf, 'Noiseless coding of correlated information sources', *IEEE Transactions on Information theory*, vol. 19, no. 4, pp. 471–480, 1973.
- [5] A. D. Wyner and J. Ziv, 'The rate-distortion function for source coding with side information at the decoder', *IEEE Transactions on Information Theory*, vol. 22, no. 1, pp. 1–10, 1976.
- [6] X. Artigas, J. Ascenso, M. Dalai, S. Klomp, D. Kubasov, and M. Ouaret, 'The DISCOVER codec: architecture, techniques and evaluation', in *Picture Coding Symposium*, 2007.
- [7] Huynh Van Luong, L. L. Raket, and S. Forchhammer, 'Re-estimation of motion and reconstruction for distributed video coding', *IEEE Transactions on Image Processing*, vol. 23, no. 7, pp. 2804–2819, Jul. 2014.
- [8] A. Abou-Elailah, F. Dufaux, J. Farah, M. Cagnazzo, A. Srivastava, and B. Pesquet-Popescu, 'Fusion of global and local motion estimation using foreground objects for distributed video coding', *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 25, no. 6, pp. 973–987, 2015.
- [9] Y. C. Shen, H. P. Cheng, J. C. Luo, Y. H. Lin, and J. L. Wu, 'Efficient real-time distributed video coding by parallel progressive side information regeneration', *IEEE Sensors Journal*, vol. 17, no. 6, pp. 1872–1883, Mar. 2017.
- [10] A. Aaron, S. D. Rane, E. Setton, and B. Girod, 'Transform-domain Wyner-Ziv codec for video', in *Electronic Imaging*, 2004, pp. 520–528.
- [11] A. Aaron, R. Zhang, and B. Girod, 'Wyner-Ziv coding of motion video', in *36th Asilomar Conference on Signals, Systems and Computers*, 2002, vol. 1, pp. 240–244.
- [12] X. Guo, Y. Lu, F. Wu, and W. Gao, 'Distributed video coding using wavelet', in *IEEE International Symposium on Circuits and Systems*, 2006, p. 4–pp.
- [13] C. Brites, J. Ascenso, and F. Pereira, 'Improving transform domain Wyner-Ziv video coding performance', in *IEEE International Conference on Acoustics Speech and Signal Processing*, 2006, vol. 2, pp. II–II.
- [14] A. Mahmood, L. Dooley, and P. Wong, 'Transform domain distributed video coding using larger transform blocks', to be presented at the 5th IEEE Global Conference on Signal and Information Processing, Montreal, Canada, 2017.
- [15] J. Park and B. Jeon, 'Rate-constrained region of interest coding using adaptive quantization in transform domain Wyner-Ziv video coding', *IEEE Transactions on Broadcasting*, vol. PP, no. 99, pp. 1–15, 2016.
- [16] Y. Mohammad Taheri, M. O. Ahmad, and M. N. S. Swamy, 'A joint correlation noise estimation and decoding algorithm for distributed video coding', *Multimedia Tools and Applications*, Apr. 2017.
- [17] D. Varodayan, A. Aaron, and B. Girod, 'Rate-adaptive codes for distributed source coding', *Signal Processing*, vol. 86, no. 11, pp. 3123–3130, Nov. 2006.
- [18] A. B. Watson, 'DCT quantization matrices visually optimized for individual images', in *Proc. SPIE*, 1993, vol. 1913, pp. 202–216.
- [19] D. Kubasov, K. Lajnef, and C. Guillemot, 'A hybrid encoder/decoder rate control for Wyner-Ziv video coding with a feedback channel', in *IEEE 9th Workshop on Multimedia Signal Processing*, 2007, pp. 251–254.
- [20] J. Slowack, J. Škorupa, N. Deligiannis, P. Lambert, A. Munteanu, and R. Van de Walle, 'Distributed video coding with feedback channel constraints', *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 22, no. 7, pp. 1014–1026, 2012.
- [21] D. Kubasov, J. Nayak, and C. Guillemot, 'Optimal reconstruction in Wyner-Ziv video coding with multiple side information', in *IEEE 9th Workshop on Multimedia Signal Processing*, 2007, pp. 183–186.
- [22] G. K. Wallace, 'The JPEG still picture compression standard', *IEEE transactions on Consumer Electronics*, vol. 38, no. 1, pp. xviii–xxiv, 1992.
- [23] C.-Y. Wang, S.-M. Lee, and L.-W. Chang, 'Designing JPEG quantization tables based on human visual system', *Signal Processing: Image Communication*, vol. 16, no. 5, pp. 501–506, 2001.
- [24] G. Petrazzuoli, M. Cagnazzo, and B. Pesquet-Popescu, 'High order motion interpolation for side information improvement in DVC', in *IEEE International Conference on Acoustics Speech and Signal Processing*, 2010, pp. 2342–2345.